**SDC 3 Appendix: Background information and formulas for the p-values and confidence interval for the problem of comparing two means.**

Brief summary: The purpose of this technical appendix is to present the detailed formulas used to compute the p-values and the confidence interval for the problem of comparing two means. The derivation of the formulas relating  and  to p-values and the derivation of the confidence interval are also presented..

The model we assume is thatare independent and normally distributed with mean  and variance  and, independently,  are independent and normally distributed with mean and variance . The problem is to make inferences about the difference in means.

The sample mean of sample  is  (presented in L62 and L93), and the sample variance of sample  is , (the sample standard deviation  is presented in L63 and L94). The standard error is  (presented in L64 and L95) and the approximate degrees of freedom is  (presented in L65 and L96). Under the assumed model, the sampling distribution of  is approximated by the Student *t* distribution with  degrees of freedom. In the equal variance case (), the sampling distribution of  where , is exactly the Student *t* distribution with  degrees of freedom.

Let  be the distribution function of the Student *t* distribution with  degrees of freedom and let  denote a random variable with this distribution. The Student t distribution is symmetric so  and  have the same distribution. The one-sided p-value for testing the null hypothesis that  against the alternative that  is

 

given by [2]. Thus  can be interpreted as the one-sided p-value for testing the null hypothesis that  against the alternative that  (i.e. the probability under the null hypothesis  that we observe a t-ratio greater than or equal to the observed value of the t-ratio ). Similarly,  can be interpreted as the one-sided p-value for testing the null hypothesis that  against the alternative that . The derivation is

 

The confidence interval [1] for  is obtained as



The critical value  , where  is the inverse of the distribution function of the Student *t* distribution.